

STATISTICS

1. (a)
2. (a)
3. (b)
4. (d)
5. (c)
6. (b)
7. (b)
8. (a)
9. (c)

10. NORMAL DISTRIBUTION

A normal distribution is the bell-shaped symmetrical frequency distribution curve of a continuous random variable. It is meant for ideal conditions.

For a data which is normally distributed,

Mean = Median = Mode = 0 and Standard deviation = -1 or +1

11. HANDLING MISSING DATA

In order to handle the missing data, we can use some imputation techniques. Imputation is the process of substituting an estimate for missing values and analysing the entire data set as if the imputed values were the true observed values.

We can use multiple imputation technique which is a flexible tool for handling missing data.

MULTIPLE IMPUTATION TECHNIQUE

Multiple imputation fills in missing values by generating plausible numbers derived from distributions of and relationships among observed variables in the data set.³ Multiple imputation differs from single imputation methods because missing data are filled in many times, with many different plausible values estimated for each missing value. Using multiple plausible values provides a quantification of the uncertainty in estimating what the missing values might be, avoiding creating false precision (as can happen with single imputation). Multiple imputation provides accurate estimates of quantities or associations of interest, such as treatment effects in randomized trials, sample means of specific variables, correlations between 2 variables, as well as the related variances. In doing so, it reduces the chance of false-positive or false-negative conclusions.

12. A/B TESTING

A/B tests consist of a randomized experiment that usually involves two variants (A and B), although the concept can be also extended to multiple variants of the same variable. It includes application of statistical hypothesis testing or "two-sample hypothesis testing" as used in the field of statistics. A/B testing is a way to compare multiple versions of a single variable, for example by testing a subject's response to variant A against variant B, and determining which of the variants is more effective.

13. MEAN IMPUTATION

Mean imputation is typically considered terrible practice since it ignores feature correlation. Consider the following scenario: we have a table with age and fitness scores, and an eight-year-old has a missing fitness score. If we average the fitness scores of people between the ages of 15 and 80, the eighty-year-old will appear to have a significantly greater fitness level than he actually does.

Also, mean imputation decreases the variance of our data while increasing bias. As a result of the reduced variance, the model is less accurate and the confidence interval is narrower.

14. LINEAR REGRESSION IN STATISTICS

What is linear regression? Linear regression analysis is used to predict the value of a variable based on the value of another variable. The variable you want to predict is called the dependent variable. The variable you are using to predict the other variable's value is called the independent variable.

For example, it can be used to quantify the relative impacts of age, gender, and diet (the predictor variables) on height (the outcome variable).

15. BRANCHES OF STATISTICS

We can split statistics into two main branches

1. Descriptive Statistics
2. Inferential Statistics

Descriptive Statistics

a. Central Tendency

Mean
Median
Mode

b. Dispersion of data

Range
Variance
Standard deviation
Percentile
Skew

Inferential Statistics

- a. Z Score Test
- b. Hypothesis Test
 - T Test
 - Chi square Test
 - Regression Test
 - ANOVA Test